

# Goal setting and attainment in Alzheimer's disease patients treated with donepezil

K Rockwood, J E Graham, S Fay, for the ACADIE Investigators

See editorial commentary, pages 471–2

*J Neurol Neurosurg Psychiatry* 2002;**73**:500–507

See end of article for authors' affiliations

Correspondence to:  
Dr K Rockwood, Queen Elizabeth II Health Sciences Centre, Geriatric Medicine Research Unit, Room 1421, 5955 Veterans' Memorial Lane, Halifax, Nova Scotia, Canada B3H 2E1;  
Kenneth.Rockwood@Dal.ca

Received 3 April 2002  
In revised form 9 July 2002  
Accepted 24 July 2002

**Objectives:** To understand the treatment goals of Alzheimer's disease (AD) patients, carers, and physicians; to estimate whether clinically important goals are met during treatment with donepezil; and to compare a measure of goal attainment with standard measures used to evaluate AD treatment.

**Methods:** In a 12 month phase IV trial, 108 patients with mild to moderate AD, their primary carers, and treating physicians set goals assigned to five domains, using Goal Attainment Scaling (GAS) as the primary outcome. Goal attainment was assessed quarterly. GAS scores were correlated with standard outcomes, including the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-cog), and the Clinician's Interview-Based Impression of Change-Plus (CIBIC-plus).

**Results:** Physicians set fewer goals (342, mean (SD) per patient=3 (1)) than patients/carers (855, mean=9 (3)), particularly in leisure (20% by physicians compared with 76% by patients/carers), and social interaction (24% versus 49%). Physicians observed statistically significant improvement in global goal attainment for six months, and patients/carers for nine months. Patients/carers described consistent goal attainment, whereas physicians observed variable effects, such as decline in cognition but improved social interaction and behaviour. Physician global GAS scores correlated highly with the CIBIC-plus at weeks 12 ( $r=-0.82$ ) and 52 ( $r=-0.80$ ), but not with the ADAS-cog ( $r=0.12$  and  $r=-0.45$ , respectively). Patient/carer global GAS scores correlated moderately with the physician's CIBIC-plus (week 12  $r=-0.51$ ; week 52  $r=-0.56$ ), and nominally with the ADAS-cog.

**Conclusions:** Patients/carers and physicians differ in their expectations and impressions of treatment effects. Clinically important changes correlated only modestly with psychometric tests. Attainment of treatment goals does not accord with a simplistic model in which successful AD treatment means that all declines uniformly improve.

Current pharmacological treatment of Alzheimer's disease (AD) chiefly consists of inhibition of the enzyme acetylcholinesterase (AChE), to result in greater availability of the neurotransmitter acetylcholine (ACh). While the "better memory through ACh" hypothesis has its critics<sup>1</sup> AChE inhibition (AChEI) has produced statistically significant differences, favouring treatment, across a range of trials.<sup>2</sup> Nevertheless, whether these differences translate into observable treatment effects in patients' daily lives is often questioned.<sup>3–8</sup>

Although no standard method of determining clinical importance is accepted, some requirements for antidementia trials have been proposed, and many seem to have been met.<sup>9</sup> A biologically plausible hypothesis exists<sup>2</sup>; there is a reproducible, dose response effect, and the outcome measures converge.<sup>9</sup> Still, a ready summary of clinical meaningfulness has proved elusive<sup>10</sup> despite the requirement that trials undertaken for regulatory approval use a global clinical measure as a primary outcome.<sup>11</sup> Of note, none of these measures incorporate the preferences of patients and carers in the assessment of benefit, even though this is held to be both an essential part of a clinically meaningful effect<sup>12</sup> and a practical guide to prescribing cholinesterase inhibitors.<sup>13</sup> Moreover, most standard measures assume that successful treatment will reverse the pattern of decline, such that regaining a certain score on a given scale in which points are lost without treatment can be used as the basis for inferences about treatment effects.<sup>5, 7, 9, 10</sup>

Clinical importance is more likely if treatment meets the expectations of patients, carers and physicians.<sup>14–16</sup> Goal Attainment Scaling (GAS) is a formal method of assaying such expectations. GAS requires that individualised treatment

goals be defined at the outset, monitored regularly, and summarised according to a formula, which classifies the degree of goal attainment. To better understand clinical meaningfulness in AD treatment, we elicited goals from patients and their carers and from treating physicians in a multicentre trial of donepezil, compared each group's goal attainment with standard assessments of treatment effects, and explored patterns of goal setting by patients/carers and physicians.

## METHODS

### Patients and carers

The Atlantic Canada Alzheimer's Disease Investigation of Expectations (ACADIE) study targeted community dwelling Alzheimer's patients judged to have dementia of mild to moderate severity according to standard criteria.<sup>17, 18</sup> Eligible patients, together with their primary carers, were recruited at six study centres across Atlantic Canada. Calculations based on a summary of donepezil efficacy results on the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-cog),<sup>19</sup> with a standard table,<sup>20</sup> and assuming  $\alpha=0.05$  in a two tailed test were used to determine that a sample size of 90 would allow 80% power to detect an effect size of 0.40. Allowing for 10% attrition, recruitment of 100 patients was planned.

**Abbreviations:** AD, Alzheimer's disease; GAS, Goal Attainment Scaling; ADAS-cog, Alzheimer's Disease Assessment Scale-Cognitive; CIBIC-plus, Clinician's Interview-Based Impression of Change-Plus; AChE, acetylcholinesterase; ACh, acetylcholine; AChEI, acetylcholine inhibition

**Table 1** Sample GAS follow up guide

|                                    | Goal areas   |  |  |
|------------------------------------|--|--|--|
|                                    | cognition (judgment-safety)  | cognition (memory)   | function (cooking)   |
| Much worse than baseline (-2)      | Unable to make any judgement or solve problems; Requires 24 hour supervision   | No longer recognises daughter  | Unable to assist in any meal preparation   |
| Somewhat worse than baseline (-1)  | Severe impairment in judgement and problem solving;<br>Unable to operate stove, microwave, or oven (even under supervision)  | Repetitious in questioning/ speech >10 times a day;<br>No longer asks for shopping list and cannot use it, even with prompting   | Attempts to assist in meal preparation 1-2 times per week, but is unable to complete task, even with assistance  |
| Baseline status (0)                | Impaired judgement with regard to safety concerns;<br>Leaves pots to cook dry on stove;<br>Put styrofoam cup on stove to heat contents;<br>Puts unsafe items in microwave for extensive length of time;<br>Family feels he would not recall or react well in emergency such as fire* | Repetition in speech, talks non-stop and repeats same thing over and over (>5 times a day);<br>Difficulty shopping, even with list, constantly rechecks cart to see for items on list;<br>Does not remember grandchildren when they come over* | No longer cooks meals by self;<br>Assists with meal preparation daily but can only manage one task and requires direction and supervision;<br>Leaves stove burners on for indefinite periods when he attempts to cook alone* |
| Somewhat better than baseline (+1) | Only slight impairment in judgement;<br>Can operate stove, oven, and microwave with supervision;<br>Is able to recall emergency plan for fire†   | Less repetitious (<5 times a day)†   | Completes 2 or 3 of the tasks involved in meal preparation (daily) with assistance and supervision (for example, peeling and cooking vegetables, making a salad, setting the table, etc)                                     |
| Much better than baseline (+2)     | Solves small everyday problems;<br>Uses good sound judgement in regards to safety issues, such as stove/oven operation and locking doors;<br>Remembers emergency plan for fire   | Able to use and remember items chosen from the list (does not re-check cart for items)   | Completes 2 or 3 of the tasks involved in meal preparation (daily) without supervision or assistance†  |

\*Initial score. †Final score.

### Design and instruments

Participants were recruited to participate in a 52 week open label, prospective study of treatment with donepezil hydrochloride. At baseline, after obtaining informed consent, diagnostic verification and standard psychometric testing, patients were given 5 mg donepezil tablets, to be taken once daily, for the first 12 weeks of the study. Follow up assessments were conducted at 12, 24, 36, and 52 weeks. If clinically indicated, the donepezil dose was increased to a daily maximum of 10 mg at (or subsequent to) the 12 week follow up visit. At each time point, patients and carers underwent a battery of standard clinical assessments and participated in an open-ended home interview conducted by trained field researchers (masters or doctoral level social scientists). All home visits were audio recorded. The findings of the clinicians were blinded from the findings of the field researchers, and vice versa.

GAS<sup>14, 16, 21-27</sup> was the primary outcome measure. Like the CIBIC-plus, GAS is a clinimetric measure (that is, one that is based on clinical judgments about individual patients) designed to capture individual treatment effects. Briefly, GAS uses a five point scale, which is individualised according to patients' needs. The scale is anchored at a baseline value, which represents the pretreatment status for individual patient problems. The scale extremities describe probable better and worse outcomes.

The GAS process begins with the selection of individualised problem areas, typically elicited by a phrase such as "Please tell me what it is like to have Alzheimer's disease. How is it a problem for you?" Each problem needing treatment is then described in observable terms and recorded on the GAS follow up guide at the 0 level (the baseline status). Next, anticipated outcome levels are set. Expectations for possible improvements are designated on the GAS follow up guide as +1, an outcome that is a little better than baseline, and +2, an outcome that is a lot better than baseline. Similarly, potential declines are specified. A little worse than baseline is recorded as -1 and a lot worse than baseline is recorded as -2. As patients are seen subsequently, their follow up status is reviewed and scored. Patients' goal attainment (that is, their score on each goal) is determined by the extent to which their status at follow up conforms to one of the predetermined levels (where 0=no change).

In this study, we set two groups of goals. The treating physicians and nurses set "clinician goals" during the baseline office visits and recorded them using their own words, based upon patient examinations, clinical observations, and information provided by the patient and carer informants. Field researchers conducted in depth home interviews with patients and carers within one week of the baseline clinic visit. During home visits, patients and their carers discussed their experiences of AD, its effect on their daily lives, and their expectations for treatment. Field researchers assisted patients and carers in the identification of treatment goals. This resulted in the development of separate "patient/carer goals" recorded by the field researcher using the patients' and carers' own words. In each case, patients, carers, and physicians could choose to set as many or as few goals as they wished. The clinical team met after each follow up visit to review and score the goals set at baseline. Field researchers saw patients and their carers at home within one week of each clinic visit to review and score the GAS follow up guides.

Based on our earlier experience,<sup>14</sup> we specified that inquiries be made, but not limited to, four general domains: cognition, function, behaviour, and leisure. Within domains, patients/carers and physicians could choose to set as many goals or as few (including none) as they wished (that is, it was not a requirement of the study that patients have a goal set in each domain). Individualised goals were categorised into domains at the study centre, where social interaction emerged as a fifth domain.

For example, a typical goal such as memory fell into the domain of cognition. A description of a memory problem might be "Dad asks me the same question over and over—more than five times a day on most days. He has trouble shopping now, even with a list because he has problems remembering whether he has picked up the items and must constantly recheck the cart. He often doesn't remember the names of the grandchildren when they come over." Using this as the baseline level of performance (scored as "0" on the GAS follow up guide, table 1), descriptions of better and worse outcomes were then elicited from the patient and/or carer. A worse outcome might be that the patient repeats questions more than 10 times per day and/or that he cannot remember

**Table 2** Participant characteristics at baseline (n=100)

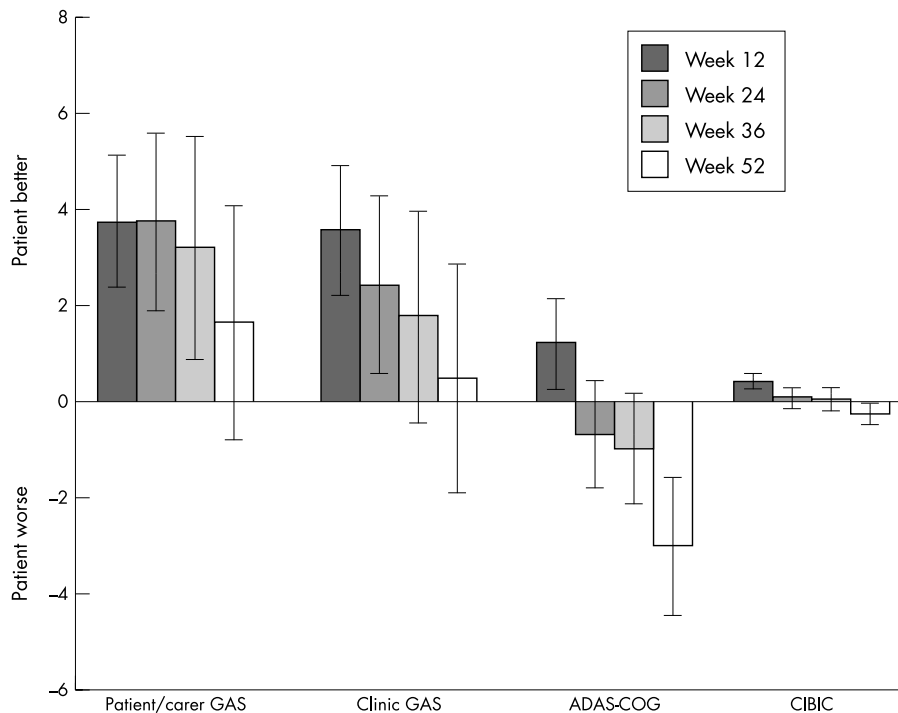
| Patient characteristics                    |                 |
|--|-----------------|
| Mean (SD) age (y) (range)                  | 76 (8) (51–96)  |
| Sex = female (%)                           | 71 (71)         |
| Mean (SD) years of education (range)       | 11 (3) (4–20)   |
| CDR Memory Score = mild (%)                | 75 (75)         |
| AD type = probable (%)                     | 98 (98)         |
| Mean (SD) duration of illness (range)      | 1.4 (1.6) (0–9) |
| MMSE (SD) (range)                          | 20 (5) (2–29)   |
| ADAS-Cog (SD) (range)                      | 25 (10) (8–58)  |
| PSMS (SD) (range)                          | 9 (3) (6–16)    |
| IADL (SD) (range)                          | 20 (5) (8–30)   |
| FAQ (SD) (range)                           | 18 (7) (2–30)   |
| Carer characteristics                      |                 |
| Mean (SD) age (y) (range)                  | 61 (14) (28–91) |
| Sex = female (%)                           | 66 (66)         |
| Mean (SD) years of education (range)       | 12 (3) (3–21)   |
| Relationship to patient = spouse (%)       | 48 (48)         |
| Resides with patient (%)                   | 66 (66)         |
| Currently employed (%)                     | 36 (36)         |
| Mean (SD) Depression Score (CES-D) (range) | 9 (8) (0–40)    |

to ask for the list and can no longer use it with prompting; a still worse outcome would be that he would no longer recognise his daughter, who is his major carer. Better outcomes would be that he asked questions less frequently, that he is able to use a list, and no longer needs to check the cart. From this inventory, five states can be demonstrated: the present state, and states somewhat and much better/worse, respectively.

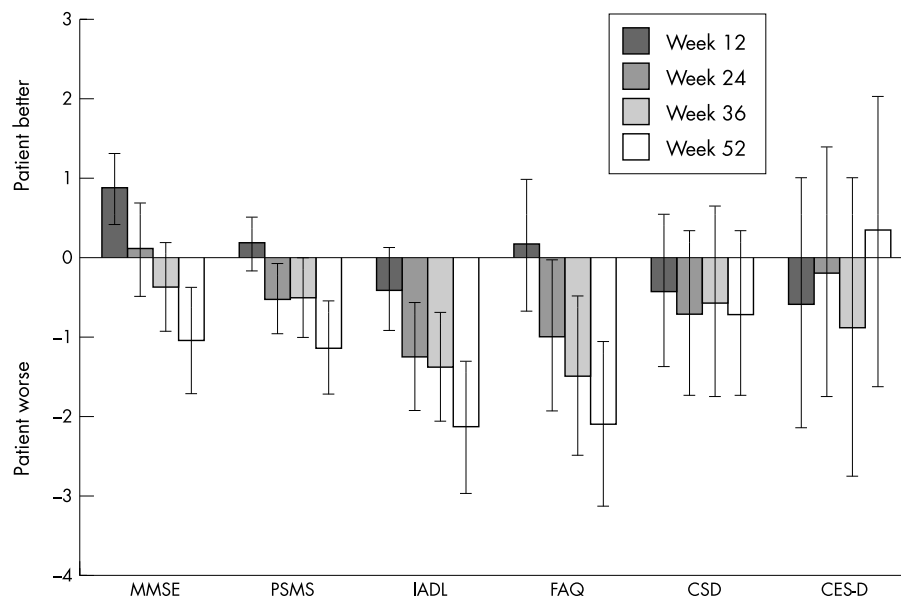
Patient/carers goals identified at baseline were also weighted on a visual analogue scale in order of relative importance. This step is optional; clinician goals were not weighted (all goals were set with weight=1). Note that goal attainment is scored by the terms specified at the outset. A global score, summarising attainment across a set of goals, is calculated as  $50 + \{ [10\epsilon(w_i x_i)] / [0.7\epsilon w_i^2 + 0.3(\epsilon w_i)^2]^{1/2} \}$ , where  $x_i$ =the individual attainment level, and  $w_i$ =the weight of that level. As each of the attainment levels is 0, so is their product, and thus the baseline score is 50. This is a mathematically equivalent modification of the conventional method,<sup>28</sup> which sets the

baseline status at “–1,” and has the “0” level represent the treatment goal. This modification incorporates more levels of decline, which seemed preferable, given that deterioration is likely.

In addition to GAS, other standardised measures were used. Cognition was assessed using the Mini-Mental State Examination (MMSE)<sup>29</sup> and the Alzheimer’s Disease Assessment Scale-Cognitive (ADAS-cog).<sup>19,30</sup> To measure various aspects of physical function we used the Physical Self-Maintenance Scale (PSMS),<sup>31</sup> Instrumental Activities of Daily Living (IADL)<sup>31</sup> and the Functional Activities Questionnaire (FAQ).<sup>32,33</sup> Depression in patients was measured using the Cornell Depression Scale (CDS)<sup>34</sup> and the Centre for Epidemiologic Studies Depression Scale (CES-D)<sup>35,36</sup> was used for carers. The Clinician’s Interview-Based Impression of Change-Plus caregiver input (CIBIC-plus)<sup>37</sup> was administered by the treating physician. The CIBIC-plus is a seven point Likert scale of change from baseline.



**Figure 1** Mean change (95% CI) from baseline for GAS, ADAS-cog, and CIBIC-plus



**Figure 2** Mean change (95% CI) from baseline for secondary efficacy measures.

**Analyses**

The analyses of treatment effects used the intention to treat last observation carried forward (ITT-LOCF) principle. Mean change and 95% confidence intervals from baseline were calculated at each follow up visit for all outcome measures. A one way repeated measures analysis of variance was performed on the change scores as a function of time for the primary outcome measures (patient/carer GAS and clinician GAS) with the Bonferonni adjustment for multiple comparisons. Otherwise the mean change values and their 95% confidence intervals are presented. To assess construct validity, Spearman correlation coefficients were calculated for GAS change scores and compared with the secondary efficacy measures; small to moderate correlations were expected. Statistical tests were interpreted at the 5% significance level.

All subjects provided written informed consent. This project was approved by the Research Ethics Committee of the Queen Elizabeth II Health Sciences Centre, Halifax, Nova Scotia, and by comparable committees of the other investigating sites.

**RESULTS**

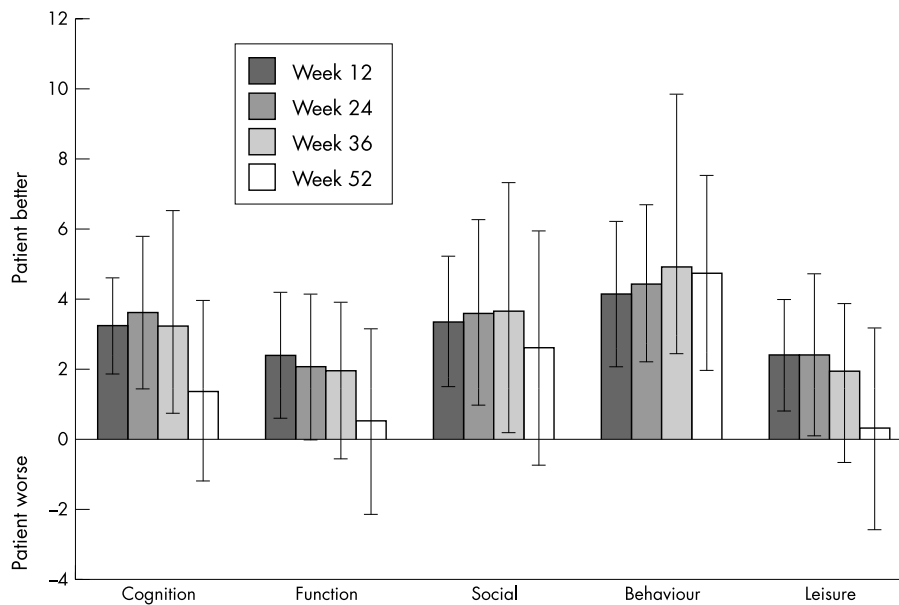
**Enrolment**

Of 108 patients enrolled, 88 (81%) successfully completed 52 weeks of treatment with donepezil. Premature discontinuation occurred because of: adverse events (n=10), withdrawal of consent (n=5), insufficient clinical response (n=2), death (n=1), and other (n=2). Of the 108 patients enrolled, eight had no observation after first treatment, leaving 100 for analysis.

**Table 3** Spearman correlation coefficients by GAS domain

|                            | Patient/Carer |         |         |         | Clinician |         |         |         |
|----------------------------|---------------|---------|---------|---------|-----------|---------|---------|---------|
|                            | Week 12       | Week 24 | Week 32 | Week 52 | Week 12   | Week 24 | Week 32 | Week 52 |
| <b>Global goals</b>        |               |         |         |         |           |         |         |         |
| MMSE*                      | 0.06          | 0.24    | 0.12    | 0.23    | 0.20      | 0.37    | 0.35    | 0.43    |
| ADAS-cog                   | 0.02          | -0.14   | -0.15   | -0.22   | 0.12      | -0.19   | -0.37   | -0.45   |
| CIBIC                      | -0.51         | -0.62   | -0.55   | -0.56   | -0.82     | -0.76   | -0.79   | -0.80   |
| PSMS                       | -0.26         | -0.43   | -0.39   | -0.51   | -0.36     | -0.42   | -0.50   | -0.54   |
| IADL                       | -0.26         | -0.27   | -0.31   | -0.45   | -0.16     | -0.39   | -0.33   | -0.44   |
| FAQ                        | -0.31         | -0.22   | -0.29   | -0.25   | -0.20     | -0.34   | -0.49   | -0.50   |
| CSD                        | -0.18         | -0.13   | -0.17   | -0.21   | -0.40     | -0.34   | -0.35   | -0.40   |
| Clinician global goals     | 0.49          | 0.53    | 0.58    | 0.51    | -         | -       | -       | -       |
| <b>Cognition</b>           |               |         |         |         |           |         |         |         |
| MMSE                       | 0.13          | 0.20    | 0.11    | 0.25    | 0.21      | 0.42    | 0.37    | 0.51    |
| ADAS-cog                   | 0.06          | -0.20   | -0.18   | -0.11   | 0.13      | -0.19   | -0.34   | -0.43   |
| Clinician cognition goals  | 0.39          | 0.56    | 0.48    | 0.41    | -         | -       | -       | -       |
| <b>Function</b>            |               |         |         |         |           |         |         |         |
| PSMS                       | -0.22         | -0.34   | -0.33   | -0.47   | -0.25     | -0.47   | -0.42   | -0.53   |
| IADL                       | -0.18         | -0.26   | -0.28   | -0.44   | -0.16     | -0.39   | -0.39   | -0.36   |
| FAQ                        | -0.10         | -0.15   | -0.21   | -0.14   | -0.17     | -0.32   | -0.47   | -0.39   |
| Clinician function goals   | 0.46          | 0.34    | 0.41    | 0.50    | -         | -       | -       | -       |
| <b>Behaviour</b>           |               |         |         |         |           |         |         |         |
| CSD-mood related signs     | 0.06          | 0.05    | -0.15   | -0.06   | -0.16     | -0.04   | -0.11   | -0.04   |
| CSD-behaviour disturbances | -0.08         | -0.02   | -0.06   | -0.05   | -0.21     | -0.16   | -0.16   | -0.26   |
| CSD-cyclic functions       | -0.11         | -0.01   | 0.14    | 0.14    | -0.21     | -0.17   | -0.18   | -0.23   |
| CSD-ideational disturbance | -0.12         | 0.01    | -0.21   | -0.10   | -0.30     | -0.47   | -0.45   | -0.23   |
| Clinician behaviour goals  | 0.41          | 0.15    | 0.33    | 0.28    | -         | -       | -       | -       |

\*In the MMSE, like GAS, a higher score indicates better performance. Thus, the expected direction is positive, whereas it is negative for the other measures.



**Figure 3** Mean change (95% CI) from baseline by domain for patient/carer GAS.

Most patients were elderly women with mild AD (table 2). Eighty two patients had treatment increased from 5 mg/day to 10 mg/day for at least one dose. At week 12, 69% of patients were receiving 10 mg/day; 67% at week 24; 66% at week 36; and 65% at week 52. Most carers were women (66%), spouses (48%), and younger (mean age=61 years) than the patient set (table 2).

**Goals and their attainment**

Patients and carers identified a total of 855 goals at baseline (mean (SD)=9 (3) per patient), while clinicians identified 342 (mean=3 (1)). Patients/carers set function goals most often (86% of patients), followed by cognition (83%), leisure (76%), behaviour (58%), and social interaction (49%). Clinicians set more cognition goals (85%), followed by function (68%), behaviour (57%), social interaction (24%), and leisure (20%).

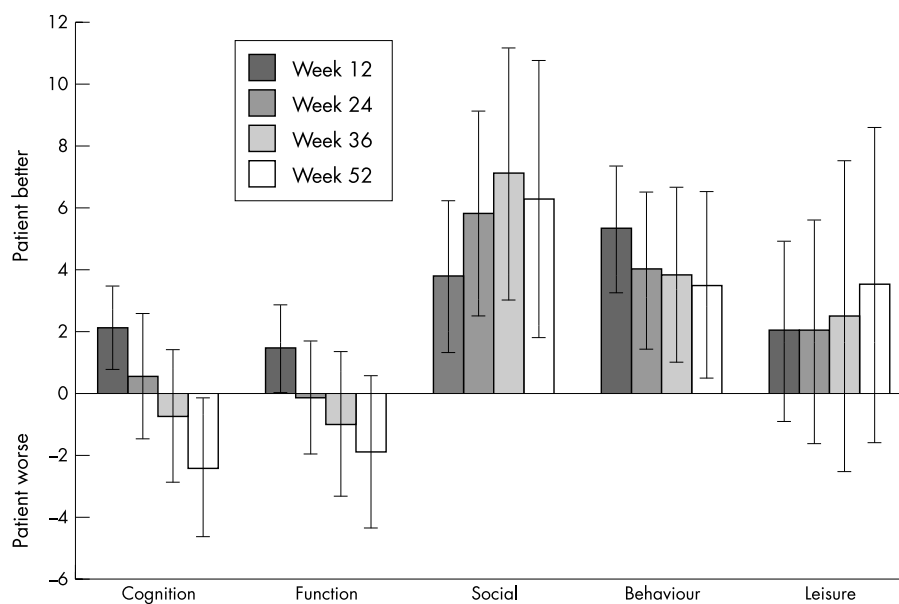
Statistically significant improvements in the global patient/carer GAS scores (fig 1) were seen to week 36 (mean change=3.19, p=0.03); however, by week 52 there was no significant difference from baseline (mean change=1.62, p=0.74). The global GAS score for clinician identified goals

improved significantly from baseline to week 24 (mean change=2.39, p=0.04), but was not significantly different from baseline thereafter (week 52 mean change=0.43, p=1.00).

**Comparisons with standard measures**

While improvement in global GAS scores was observed as noted, the standard cognitive measures (figs 1 and 2) showed significant improvements only at week 12 (MMSE mean change=0.86 and ADAS-Cog mean change=-1.17). Decline from baseline was observed for both measures thereafter (for example, week 52: mean change=-1.04 and mean change =3.07, respectively). The functional assessments showed patterns of initial maintenance of functional performance, followed by later decline (fig 2), with slight differences by level of function (chiefly IADL decline after weeks 24-36, whereas ADL function was more stable). Neither patient (CSD) or carer (CES-D) measures of depression demonstrated much change (fig 2).

Clinician global GAS scores were highly correlated (table 3) with the CIBIC-plus at weeks 12 and 52 but not with the



**Figure 4** Mean change (95% CI) from baseline by domain for clinic GAS.

ADAS-cog results. Patient/carer global GAS scores were only moderately correlated with the CIBIC-plus. Low correlations were observed between patient/carer GAS and the ADAS-cog.

Low correlations were observed between the GAS domains of cognition, function, and behaviour and corresponding secondary efficacy measures at week 12, for both clinicians and patient/carers. Moderate correlations appeared between clinician assessed GAS cognition goals and the MMSE ( $r=0.51$ ) and the ADAS-cog ( $r=-0.43$ ) at week 52. Moderate correlations were also observed between clinician function goals and the PSMS ( $r=-0.53$ ), patient/carer function goals and the PSMS ( $r=-0.47$ ), and patient/carer function goals and IADL scores ( $r=-0.44$ ). All remaining correlation comparisons at week 52 were low.

### Patterns of goal attainment, by domain

Within goal areas (fig 3), patients and carers generally observed improvement, and each domain tended to agree with the others, and thus with the global GAS score. The largest and greatest treatment effects were observed in behaviour.

The pattern of treatment effects observed by physicians was somewhat more complex. They recorded behaviour and social interaction goals above baseline at each time point (fig 4) but in contrast with the patient/carer goals, goal attainment did not proceed in the same direction for each domain. Clinician identified leisure goals changed little, on average. Similar to the standard cognitive measures, clinician identified cognition and function goals had been met at week 12 but these initial gains were not maintained.

### Adverse events

Of the 108 patients, 96% ( $n=104$ ) experienced at least one adverse event. The most frequently reported were: pain ( $n=27$ ), diarrhoea ( $n=24$ ), nausea ( $n=23$ ), dizziness ( $n=20$ ), headache ( $n=19$ ), dyspepsia ( $n=15$ ), and influenza-like symptoms ( $n=15$ ). Ten patients withdrew because of adverse events, which included: anxiety ( $n=2$ ), weight loss ( $n=2$ ), diarrhoea ( $n=1$ ), pacing ( $n=1$ ), transient ischaemic attack ( $n=1$ ), confusion ( $n=1$ ), agitated depression ( $n=1$ ), and foot pain ( $n=1$ ). Twenty two serious adverse events were reported in 16 patients, one of who died of a myocardial infarction. No serious adverse events occurred more than once and none had a clear relation to the study drug.

## DISCUSSION

In this 12 month, open label study, treatment with donepezil resulted in clinically detectable effects in patients with mild to moderate AD, which largely met the goals of patients, carers, and clinicians for six months. Thereafter, treatment accorded with expectations to varying degrees. The extent of goal attainment was not reflected in standard measures used to evaluate antidementia drugs.

Patients and carers set more goals than physicians, perhaps a product of the more open-ended nature of the field researcher interviews. Physicians seemed to lump items that patients and carers tended to split. For example, whereas patients and carers might describe goals with respect to repetitive questions, needing lists and having to have help in following routines, physicians typically would group all of these under a single goal of memory. In addition, patients and carers set more goals in the domains of leisure, social interaction, behaviour, and function, and fewer in cognition. This probably reflects competing understandings of the disease. Physicians see patients episodically and they seem to think of pathophysiological processes and manifestations that they can test, or about which they are used to inquiring.<sup>38</sup> These differences, however, do suggest that physicians might wish to explore these other areas more systematically if they are not to miss important treatment effects. At the same time, physicians are able to offer some objectivity. They can help patients

and carers sort out patterns in their observations, and not be unduly influenced by a single memorable event or by incidents that are not representative but whose saliency is raised by having happened recently. This objectivity, however, must be to some extent traded off with physicians' more episodic interactions, and their differing priorities in understanding the disease and its treatment effects. Through qualitative analyses we plan to explore these differences more systematically.

It seems, however, that as each party brings information to bear, physicians can provide a useful role by collaborating with patients and carers in understanding disease treatment. Additional qualitative analyses will allow us to comment more fully on these differences. We plan to investigate further what we see as another useful finding of this study: attainment of treatment goals does not accord with a simplistic model in which successful treatment is interpreted as meaning that all declines are recovered. For example, the analysis of the clinician goal attainment scores by domain shows that treatment, which overall rated a clinical impression of improvement, nevertheless encompassed areas in which no clinically important change was seen, and even others in which decline occurred. This accords with earlier observations that suggest that judgement about treatment effects trade off between domains.<sup>11-14</sup> Understanding such trade offs, providing a role for patient and carer involvement, and clarifying treatment expectations will be necessary if the recommendation that a global clinical measure be used in routine monitoring of cholinesterase inhibitor use in AD is taken up.<sup>10</sup>

Whether such trade offs account for the at best modest correlations between GAS scores and the scores on standard tests is not clear. In this study, with the exception of the CIBIC-plus, most correlations between other measures and GAS were in the low to moderate range, although they tended to increase over time. By week 52, the clinician GAS correlation with other measures ranged from 0.40–0.54, and the patient/carer GAS correlated between 0.21–0.51 with standard measures. How to interpret this range is not clear. On the one hand, no correlation might mean that the measure either is not valid, or is tapping entirely different information. On the other hand, high correlation might mean that the new measure adds little value to existing information. In earlier studies, we have seen similar correlations, which we have interpreted as supporting validity, but adding new information.<sup>25-26</sup> Still, none of the usual measures represents a so called "gold standard", and their clinical importance is suspect. In consequence, we see that replication offers the best guide to more completely understanding the validity of this approach.

Our study is subject to important limitations. It is not a controlled trial. Our question was not whether donepezil is more effective than placebo over 3, 6, 9, or 12 months, but rather whether treatment of AD meets a priori expectations. Our understanding is that, up to six months, the question of whether cholinesterase inhibitors provide statistically significant change in cognitive measures under double blind conditions has been answered<sup>5-7, 39-41</sup> and new data seem to extend this time frame to about 12 months.<sup>42-43</sup> We are seeking to help translate whether the treatment effects observed in earlier studies have meaning for treating physicians, and for patients and carers. In this context, the patients enrolled in the study seem comparable to those in the treatment arms of the double blinded studies. As in an earlier clinical practice study<sup>8</sup> they were slightly older (76 (8) years) than patients from three pivotal double blinded trials (combined mean age =73 (5) years). Our study had more women (71% versus 60% overall).<sup>39-41</sup> As in other studies, most of our patients (75%) had mild AD (that is, 80% overall). Considering the ADAS-cog, we found that the mean difference in scores on individually optimal doses (that is, either 5 mg or 10 mg of donepezil) was  $-1.17$  at 12 weeks and  $0.72$  at 24 weeks. These estimates are comparable with other published reports of  $-0.67$  points (24

week/end point analysis; 5 mg) and  $-1.06$  (24 week/end point analysis; 10 mg) and about 0.35 points (24 week/end point analysis 5 mg) and about  $-1.5$  (24 week/end point analysis 10 mg) and about  $-0.2$  (12 weeks/end point analysis 5 mg) and about  $-0.3$  (12 week/end point analysis 10 mg).

Patients and their carers bring a wide range of expectations in relation to AD and treatment effects. Still, patient and carer involvement in goal setting seems both feasible and valid. For example, there were only a small number of goals (about 1%) where donepezil was unlikely, a priori, to have had a treatment effect. These included goals set for mobility, incontinence, pain, and headaches.

The scoring instructions for GAS were specified so that a problem area was identified and the current level of functioning (that is, the present problem) was described at the "0" level. In 106 patient/carer goals (12%), however, while future problems were anticipated, the present level of function was satisfactory. In such cases, goals identified by patient/carer dyads who felt that maintenance was the best that could be hoped for were scored as "0". Where no improvement could be anticipated, there is no room to capture improvement. Similarly, in the 97 goals (11%) where patient/carer dyads were unwilling to specify a worse outcome, the baseline was scored as "0" and only unspecified worsening could be captured. Overall, these effects are likely to be largely offsetting, but they do reveal the tension in trying to incorporate patient and carer preferences in the face of a neurodegenerative illness where deterioration is a likely outcome. This can influence not only the appreciation of the present state, but can sometimes reveal situations in which people, perhaps as an aspect of coping, choose not to specify how things might get worse. The neurodegenerative nature of AD also means that the ITT-LOCF strategy might be less cognitive if patients who dropped out early had observations carried forward that did not reflect subsequent decline. We therefore also conducted an observed cases analysis, which did not change any of the main effects here observed.

Our data also have some strengths. We systematically incorporated the expectations of patients and carers, and their treating physicians into an evaluation of treatment. As noted, despite the broad range of expectations that patients and carers bring to the prospect of treatment for AD, they largely set goals that reflect a realistic understanding of what might be possible. Importantly, maintenance often is explicitly understood as the best outcome of treatment. Nevertheless, many of the individual goals hope for a reversal to a previous state, suggesting that the expectations are complex. Maintenance is often desirable, but when people are confronted with describing what might be better, they are likely to specify "the way things were." The previous state strongly influences the expectation of likely future outcomes.

These data are relevant to understanding the CIBIC-plus. Like it, GAS relies on clinical judgments about individual patients with caregiver input. Recent attempts to standardise the CIBIC-plus, however, have seen increasing standardisation of the domains of inquiry. By contrast, GAS incorporates patient and carer preferences as the basis for evaluation of treatment effects. For example, the test of recent memory would not be whichever standard test a given clinician uses, which is commonly how such inquiries are operationalised using the CIBIC-plus.<sup>11</sup> Rather, the inquiry might be about repetitive questioning, recalling visits, or not having to constantly recheck a grocery list, if that is how the memory problem is seen by patients and carers. Like the CIBIC-plus, GAS probably works best when used in conjunction with, and not as a substitute for, standard measures.

To date, almost all antidementia drug trials have used standard scales and staging systems exclusively. While these allow for important insights, they do not provide an accurate portrayal of the expectations, motivations, and experiences of the patients, their primary carers, and families. In addition,

these standard instruments largely assume that as AD gets better the result is chiefly a reversal of the decline in signs and symptoms that mark the disease as it progresses. This might not be the case—what seems more likely is that patients exhibit new combinations of symptoms and signs, as some that were present at baseline improve, while others worsen and still others stay the same.<sup>11–44</sup> In consequence, we run the risk that new antidementia drugs could bring about systematic and reproducible benefits, which might be missed because our current scales are not sensitive to the new profiles experienced by patients. Our study suggests that important treatment effects might not be reflected in the standard tests now in use as outcome measures in antidementia drug trials and treatments. Judging by its range of correlations with standard measures, GAS seems to be tapping into different constructs than the usual domains of inquiry now used to evaluate treatment effectiveness. At the same time, the extent of correlation is in the range that suggests convergent construct validity of the measures. Validity is likewise suggested by the higher patient/carer-clinician GAS scores. In consequence, we suggest that GAS can supplement the CIBIC-plus (but not replace it) so as to explicitly incorporate patient preferences into our understanding of whether treatment is beneficial.

## ACKNOWLEDGEMENTS

The ACADIE Investigators are: Peter Bailey MD (Saint John, NB), Mary Gorman MD (Antigonish, NS), Reginald Hutchings MD (Charlottetown, PEI), Julie Chandler MD (Yarmouth, NS), Roger McKelvey MD (Fredericton NB), Daniel Carver MD, MSc, Christopher MacKnight MD, MSc, Laurie Mallery MD (Halifax, NS). The study was supported by a grant from Pfizer Canada. Kenneth Rockwood and Janice Graham are each supported by Investigator Awards from the Canadian Institutes of Health Research, and Kenneth Rockwood received additional career support from the Dalhousie Medical Research Foundation as the Katheryn Allen Weldon Professor of Alzheimer Research.

## Authors' affiliations

**K Rockwood, S Fay**, Department of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada  
**J E Graham**, Department of Bioethics, Dalhousie University

Competing interests: KR has received speaker fees, consulting honorariums and research support from Pfizer, the distributor of donepezil, and from Janssen-Cilag and Novartis. JG has received research funds from Pfizer Canada and the Novartis Foundation for Gerontological Research. In 1998 she received speaker's fees from Hoechst, Marion, Roussel Inc.

## REFERENCES

- 1 **Davis KL**, Mohs RC, Marin D, et al. Cholinergic markers in elderly patients with early signs of Alzheimer disease. *JAMA* 1999;**281**:1401–6.
- 2 **Francis PT**, Palmer AM, Snape M, et al. The cholinergic hypothesis of Alzheimer's disease: a review of progress. *J Neuro Neurosurg Psychiatry* 1999;**66**:137–47.
- 3 **Mayeux R**, Sao M. Treatment of Alzheimer's disease. *N Engl J Med* 1999;**341**:1670–9.
- 4 **Pryse Phillips W**. Do we have drugs for dementia? No. *Arch Neurol* 1999;**56**:735–7.
- 5 **Birks JS**, Melzer D, Beppu H. Donepezil for mild and moderate Alzheimer's disease (Cochrane Review). In: *The Cochrane Library*, Issue 1, 2001. Oxford: Update Software.
- 6 **Olin J**, Schneider L. Galantamine for Alzheimer's disease (Cochrane Review). In: *The Cochrane Library*, Issue 1, 2001. Oxford: Update Software.
- 7 **Birks J**, Grimley Evans J, Iakovidou V, et al. Rivastigmine for Alzheimer's disease (Cochrane Review). In: *The Cochrane Library*, Issue 2, 2001. Oxford: Update Software.
- 8 **Greenberg SM**, Tennis MK, Brown LB, et al. Donepezil therapy in clinical practice. *Arch Neurol* 2000;**57**:94–9.
- 9 **Rockwood K**, MacKnight C. Assessing the clinical importance of statistically significant improvement in anti dementia drug trials. *Neuroepidemiology* 2001;**20**:51–6.
- 10 **Swanwick GRJ**, Lawlor BA. Initiating and monitoring cholinesterase inhibitor treatment for Alzheimer's disease. *Int J Geriatr Psychiatry* 1999;**14**:244–8.
- 11 **Joffres C**, Graham J, Rockwood K. Qualitative analyses of the clinicians' interview-based impression of change (plus): methodological

- issues and implications for clinical research. *International Psychogeriatrics* 2000;**12**:403–13.
- 12 **Winblad B**, Brodaty H, Gauthier S, *et al*. Pharmacotherapy of Alzheimer's disease: is there a need to redefine treatment success? *Int J Geriatr Psychiatry* 2001;**16**:653–66.
  - 13 **Spilker B**. *Guide to clinical trials*. New York, NY: Raven Press, 1991.
  - 14 **Rockwood K**, Stolee P, Howard K, Mallery L. Use of goal attainment scaling to measure treatment effects in an anti-dementia drug trial. *Neuroepidemiology* 1996;**15**:330–3.
  - 15 **Bogardus ST Jr**, Bradley EH, Tinetti ME. A taxonomy for goal setting in the care of persons with dementia. *J Gen Intern Med* 1998;**13**:675–80.
  - 16 **Hartman D**, Borrie MJ, Davison E, *et al*. Use of goal attainment scaling in a dementia special care unit. *Am J Alz Dis* 1997;**12**:111–16.
  - 17 **McKhann G**, Drachman D, Folstein M, *et al*. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;**34**:939–44.
  - 18 **Clarfield AM**. Canadian Consensus Conference on the Assessment of Dementia. *Can Med Assoc J* 1991;**144** (suppl).
  - 19 **Rosen WG**, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry* 1984;**141**:1356–64.
  - 20 **Kraemer HC**, Thieman S. *How many subjects? Statistical power analysis in research*. Newbury Park: Sage, 1987.
  - 21 **Kiresuk TJ**, Sherman RE. Goal Attainment Scaling: A general method for evaluating comprehensive community mental health programs. *Community Ment Health J* 1968;**4**:443–53.
  - 22 **Stolee P**, Rockwood K, Fox RA, *et al*. The use of goal attainment scaling in a geriatric care setting. *J Am Geriatr Soc* 1992;**40**:574–8.
  - 23 **Joyce B**, Rockwood K, Mate-Kole C. Use of Goal Attainment Scaling in brain injury in a rehabilitation hospital. *Am J Phys Med Rehabil* 1994;**37**:10–14.
  - 24 **Gordon J**, Powell C, Rockwood K. Goal Attainment Scaling as a measure to assess clinically important change in the nursing home. *Age Ageing* 1999;**28**:275–81.
  - 25 **Rockwood K**, Stolee P, Fox RA. The Use of Goal Attainment Scaling in measuring clinically important change in the frail elderly. *J Clin Epidemiol* 1993;**46**:1113–18.
  - 26 **Rockwood K**, Joyce B, Stolee P. Use of Goal Attainment Scaling in measuring clinically important change in rehabilitation patients with traumatic brain injury. *J Clin Epidemiol* 1997;**50**:581–8.
  - 27 **Rockwood K**, Stodnyk K, Carver D, *et al*. A clinimetric evaluation of specialized geriatric care for frail elderly people. *J Am Geriatr Soc* 2000;**48**:1080–5.
  - 28 **Kiresuk TJ**, Smith RE, Cardillo JE. *Goal Attainment Scaling: applications theory and measurement*. Hillside, NJ: Lawrence Erlbaum Associates, 1994.
  - 29 **Folstein MF**, Folstein SE, McHugh PR. Mini-mental-state: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatric Res* 1975;**12**:189–98.
  - 30 **Standish T**, Molloy DW, Bedard M, *et al*. Improved reliability of the Standardized Alzheimer's Disease Assessment Scale (SADAS) compared with the Alzheimer's Disease Assessment Scale (ADAS). *J Am Geriatr Soc* 1996;**44**:712–16.
  - 31 **Lawton MP**, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;**9**:179–86.
  - 32 **Pfeffer RI**, Kurosaki TT, Harrah CH, *et al*. Measurement of functional activities in older adults in the community. *J Gerontol* 1982;**37**:323–9.
  - 33 **Pfeffer RI**, Kurosaki TT, Chance JM, *et al*. Use of the mental function index in older adults: reliability, validity and measurement of change over time. *Am J Epidemiol* 1984;**120**:922–35.
  - 34 **Abrams A**, Young S. Cornell scale for depression in dementia. *Biol Psychiatry* 1988;**23**:271–84.
  - 35 **Lyness JM**, Kelly Noel T, Cox C, *et al*. Screening for depression in elderly primary care patients. *Arch Intern Med* 1997;**157**:449–54.
  - 36 **Radloff LS**. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;**1**:385–401.
  - 37 **Knopman DS**, Knapp MJ, Gracon IS, *et al*. The Clinician Interview-Based Impression (CIBI): a clinician's global change rating scale in Alzheimer's disease. *Neurology* 1994;**44**:2315–21.
  - 38 **Graham JE**. *Diagnosing dementia: signs, symptoms and meaning*. [Doctoral dissertation]. Quebec: University of Montreal, 1996. National Archives of Canada.
  - 39 **Rogers SL**, Friedhoff LT, the Donepezil Study Group. The efficacy and safety of donepezil in patients with Alzheimer's disease: results of a US multicentre, randomized, double-blind, placebo-controlled trial. *Dementia* 1996;**7**:293–303.
  - 40 **Rogers SL**, Farlow MR, Doody RS, *et al*. A 24-week double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. *Neurology* 1998;**50**:136–45.
  - 41 **Burns A**, Rossor M, Hecker J, *et al*. The effects of donepezil in Alzheimer's disease - results from a multinational trial. *Dement Geriatr Cogn Disord* 1999;**10**:237–44.
  - 42 **Winblad B**, Engedal K, Soininen H, *et al*. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. *Neurology* 2001;**57**:489–95.
  - 43 **Mohs RC**, Doody RS, Morris JC, *et al*. A 1-year placebo-controlled preservation of function survival study of donepezil in AD patients. *Neurology* 2001;**54**:481–8.
  - 44 **Mitnitski AB**, Graham JE, Mogilner AJ, *et al*. The rate of decline of functions in Alzheimer's disease and other dementias. *J Gerontol Med Sci* 1999;**54A**:M65–9.